

## 2. Data File Variables and Coding Schemes

---

The rectangular data file included in the *AIDS Public Information Data Set* contains one line of data for each AIDS case reported to CDC. Each line contains 30 columns. The columns contain 16 variables extracted from CDC's national AIDS data set. Please note that some of the variables in the December 1995 data set are located in different columns than in previous versions. To allow inclusion of specific geographic information (i.e., MSA) while prohibiting indirect identification of patients, some variables included in previous versions have been deleted.

Column	Variable	Description
1	<i>age</i>	Age group at diagnosis of the first AIDS-indicator opportunistic condition
2	<i>sexclass</i>	Sexual classification of patient
3	<i>race</i>	Race of patient
4	<i>categ</i>	Indicates which of the CDC AIDS case revisions the patient meets
5-8	<i>dxdate</i>	Month of diagnosis of first AIDS-indicator opportunistic condition
9-12	<i>repxdate</i>	Date when CDC first received information about the case
13-14	<i>survival</i>	For patients reported dead, number of quarters between diagnosis and death
15-16	<i>exposure</i>	Mode of exposure to HIV
17	<i>multrisk</i>	Indicates if patient had more than one risk of exposure to HIV
18	<i>birth</i>	Country of birth
19	<i>sexbi</i>	Sex with a bisexual man (women only)
20	<i>sexiv</i>	Sex with an injecting drug user
21	<i>sexother</i>	Sex with a person with hemophilia or with a transfusion recipient
22	<i>sexhiv</i>	Sex with a person known to be infected with HIV or to have AIDS, but whose mode of exposure is unknown
23-28	<i>adjwgt</i>	Reporting delay adjustment weight
29-32	<i>msa</i>	Region of residence at diagnosis of AIDS

Each of these variables is coded alpha-numerically. The codes used in the *AIDS Public Information Data Set* are described below.

### Age (column 1)

This variable contains the patient's age when he or she was first diagnosed with an AIDS-indicator disease.

0 = Less than 1 year old

1 = 1 to 12 years old

2 = 13 to 19 years old

3 = 20 to 24 years old  
4 = 25 to 29 years old  
5 = 30 to 34 years old  
6 = 35 to 39 years old  
    or age is missing  
7 = 40 to 44 years old  
8 = 45 to 49 years old  
9 = 50 to 54 years old  
A = 55 to 59 years old  
B = 60 to 64 years old  
C = 65 years old or older

## Sexclass (column 2)

Adult/adolescent males are classified according to their sexual orientation.

1 = Adult/adolescent male who has sex only with other men or sex is missing  
2 = Adult/adolescent male who has sex with both men and women  
3 = Adult/adolescent heterosexual male or pediatric male  
4 = Female (both adult/adolescent and pediatric)

## Race (column 3)

1 = White (not Hispanic)  
2 = Black (not Hispanic)  
3 = Hispanic  
4 = Asian/Pacific Islander  
5 = American Indian/Alaskan Native  
9 = Unknown

## Categ (column 4)

This variable reflects changes made over time to the CDC surveillance definition for AIDS. Only cases meeting the current (1993) surveillance definition are included in this data set. *Categ* indicates whether the patient also met the pre-1985, 1985, or 1987 surveillance definition, and whether the diagnosis, if it meets the 1987 or 1993 definition, was definitive or presumptive. Cases that meet more than one of these surveillance definitions are classified into the category listed first. For more information about the 1993 definition, see *Morbidity and Mortality Weekly Report*, December 18, 1992, Recommendations and Reports.

- 1 = Case meets the pre-1985 surveillance definition
- 2 = Case meets the 1985 surveillance definition
- 3 = Case meets the 1987 surveillance definition and was diagnosed definitively
- 4 = Case meets the 1987 surveillance definition and was diagnosed presumptively
- 5 = Case meets the 1993 surveillance definition: pulmonary tuberculosis, recurrent pneumonia, and/or cervical cancer (definitive diagnosis)
- 6 = Case meets the 1993 surveillance definition: pulmonary tuberculosis and/or recurrent pneumonia (presumptive diagnosis)
- 7 = Case meets the 1993 surveillance definition, severe HIV-related immunosuppression

### Dxdate (columns 5 through 8)

This variable contains the year and month in which the first AIDS-indicator condition was diagnosed. Columns 5 and 6 contain the year; columns 7 and 8 contain the month. Cases diagnosed before 1982 are coded as “8199.” Cases whose month of diagnosis is unknown are coded as “99” in the month portion of this variable.

### Reptime (columns 9 through 12)

This variable contains the year and month in which CDC received the case report. Columns 9 and 10 contain the year; columns 11 and 12 contain the month. Cases reported during 1981 are coded as “8199.”

### Survival (columns 13 and 14)

For patients reported dead, *survival* contains the number of quarters between diagnosis of AIDS and date of death. This variable is blank for patients whose AIDS diagnosis was made in the most recent quarter, regardless of their vital status. Patients whose survival time is greater than 32 quarters (8 years) are coded as “33.” *Survival* is coded as “0” for patients whose report date to CDC is earlier than their date of diagnosis (see page 12). The 1993 revision to the AIDS surveillance case definition allows patients to be diagnosed with AIDS based on laboratory evidence of severe immunosuppression. For many patients, this is earlier in the course of illness than diagnosis based on AIDS-defining opportunistic conditions. Thus the survival time for these patients may be longer. Changes in survival trends, particularly around 1993, should be interpreted with caution.

### Exposure (columns 15 and 16)

For surveillance purposes, AIDS cases are counted only once in a hierarchy of exposure categories. Persons with more than one reported mode of exposure to HIV are classified in

the exposure category listed first in the hierarchy, except for men with both a history of sexual contact with other men and injecting drug use. They make up a separate exposure category. Persons with multiple reported modes of exposure are indicated in the variable *multrisk*.

“Men who have sex with men” cases include men who report sexual contact with other men (i.e., homosexual contact) and men who report sexual contact with both men and women (i.e., bisexual contact). “Heterosexual contact” cases are in persons who report specific heterosexual contact with a person with, or at increased risk for, HIV infection (e.g., an injecting drug user).

Adults/adolescents born, or who had sex with someone born, in a country where heterosexual transmission was believed to be the predominant mode of HIV transmission (formerly classified as Pattern-II countries by the World Health Organization) are no longer classified as having heterosexually acquired AIDS. Similar to case reports for other persons who are reported without behavioral or transfusion risks for HIV, these reports are now classified (in the absence of other risk information which would classify them into another exposure category) as “no risk reported or identified” (see *Morbidity and Mortality Weekly Report*, March 11, 1994). Children whose mother was born, or whose mother had sex with someone born, in a Pattern-II country are now classified (in the absence of other risk information which would classify them into another exposure category) as “Mother with/at risk for HIV infection: has HIV infection, risk not specified.”

“Risk not reported or identified” cases are in persons with no reported history of exposure to HIV through any of the routes listed in the hierarchy of exposure categories. Risk not reported or identified cases include persons who are currently under investigation by local health department officials; persons whose exposure history is incomplete because they died, declined to be interviewed, or were lost to follow-up; and persons who were interviewed or for whom other follow-up information was available and no exposure mode was identified. Persons who have an exposure mode identified at the time of follow-up are reclassified into the appropriate exposure category.

#### Adult/adolescent exposure categories

- 1 = Men who have sex with men
- 2 = Injecting drug use
- 3 = Men who have sex with men and inject drugs
- 4 = Hemophilia/coagulation disorder
- 5 = Heterosexual contact with a person with, or at increases risk for, HIV infection
- 7 = Receipt of blood transfusion, blood components, or tissue
- 8 = Risk not reported or identified

## Pediatric exposure categories

- 9 = Hemophilia/coagulation disorder
- 10 = Mother with, or at risk for, HIV infection
- 11 = Receipt of blood transfusion, blood components, or tissue
- 12 = Risk not reported or identified

## Multrisk (column 17)

*Multrisk* is coded only for adult/adolescent patients (13 years old or older) and indicates if the patient has risk(s) of exposure to HIV other than the one indicated by *exposure*.

- 0 = Patient's only mode of exposure to HIV is that indicated by *exposure*
- 1 = Patient has additional risk(s) of exposure
- 2 = Patient's mode of exposure is not reported or identified

## Birth (column 18)

- 1 = Patient was born in the United States or its dependencies and possessions, or place of birth was not specified
- 2 = Patient was born outside the United States

## Heterosexual risk information (columns 19 through 22)

These variables (*sexbi*, *sexiv*, *sexother*, and *sexhiv*) contain additional exposure information for patients infected heterosexually. All 4 variables are coded as follows:

- 0 = no
- 1 = yes
- 9 = missing/unknown

The variable *sexbi* is coded only for women (for men, the variable contains a blank). All 4 variables contain "9" (missing/unknown) for patients with hemophilia, regardless of whether the exposure information is in fact unknown. This restriction is necessary in order to comply with the Assurance of Confidentiality on page 5. Of the 4,596 AIDS cases reported through December 1995 among adults/adolescents with hemophilia, less than 4 percent also reported heterosexual contact with a person at increased risk for AIDS or HIV infection.

## Adjwgt (columns 23 through 28)

This variable contains an adjustment weight which, when used as a weighting variable in a frequency tabulation, produces tabulations of AIDS cases that are adjusted for delays in case reporting (see page 11 for a discussion of delays in reporting). The weights are based on estimated reporting delay distributions that take into account exposure, geographic, and demographic variations in case reporting. The adjustment weights and the resulting tabulations are not reliable for cases diagnosed during the most recent 6 months. The utility menu contains an adjusted weight option. If you select this option, all subsequent tabulations you request will be weighted accordingly.

## MSA (columns 29 through 32)

Metropolitan area of residence at diagnosis of AIDS is identified for adult/adolescent patients residing in MSAs with 500,000 or more population, according to the latest available official U.S. Bureau of Census estimates. Each MSA is identified by a 4-digit code listed in Appendix B. For adult/adolescent patients residing in an MSA with less than 500,000 population, in a non-metropolitan area, or whose metropolitan area of residence is unknown, and for all pediatric patients, region of residence is identified. The regional codes are:

- 1 = Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont
- 2 = Midwest: Indiana, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin
- 3 = South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia
- 4 = West: Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming
- 5 = U.S. dependencies, possessions, and independent nations in free association with the United States: Guam, Puerto Rico, the U.S. Virgin Islands, and the U.S. Pacific Islands listed on page 8.